



Figure 1

The bi-directional binary tree structure with three leaf nodes. Note that a parent node has two pointers to its child nodes and each child node has a pointer back to its parent.

GUGCACC	matchingTimes	3	
	matchedOrg	E.colik12	Ø
		B.subtilis	Ø
		Sncy.6803b	Ø
	treeNodeValues	301	1.00
		305	0.95
		402	0.83
		465	0.71
		590	0.64
UCGGGAC	matchingTimes	4	
	matchedOrg	Acp.laidla	Ø
		Cor.xerosi	Ø
		Hel.choltra2	Ø
		C.leptum	Ø
	treeNodeValues	2304	1.00
		2305	1.00
		2310	0.94
		2323	0.76
		2330	0.58

Figure 2

The elaborate structure of the composite hash used in the program. Only two entries are shown in this figure. A hash is represented by a table and the keys are shaded. Ø denotes the data type "undef" in Perl. The data in this hash are for elucidatory purposes only.


```

LOCUS      E.colirnA3  3714 bp  RNA          RNA          09-NOV-1998
DEFINITION Escherichia coli str. MG1655 [gene=rrsA gene].
REFERENCE  1
AUTHORS    Blattner,F.R., Plunkett,G.,III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,R.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
TITLE      The complete genome sequence of Escherichia coli K-12
JOURNAL    Science 277 (5331), 1453-1474 (1997)
COMMENT
            Corresponding GenBank entry: U00096 (bases 4033120 to 4034661)
            legacy_attribute= CG Site No. 189
            operon= rrsA gene
            isolate_name= MG1655
BASE COUNT 389 a 352 c 487 g 314 u 2172 others
ORIGIN

```

```

1  ~~~~~ AAAUUGA A-GAGUU-U- GA-U-CAU-G
.
3541 -GUAGG-GGA--A-CCUG--C GGU--UG-GA UCACCUCCUU A~~~~~
3601 ~~~~~
3661 ~~~~~

```

//

readseq

```

>E.colirnA3  3714 bp  RNA          RNA          09-NOV-1998, 3714 bases, 1504 checksum.
            ~~~~~ AAAUUGAA-GAGUU-U-
GA-U-CAU-G
.
GUAGG-GGA-A-CCUG--CGGU--UG-GAUCACCUCCUUA~~~~~
~~~~~
~~~~~
~~~~~

```

fasta2flat

```

E.colirnA3  AAAUUGAAGAGUUGAUCAUG..GUAGGGGAACCUCCGGUUGGAUCACCUCCUUA

```

5 Figure 4

Subsystem I converts the format of the sequence file as shown schematically above.